PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/pccp

PAPER

Structure and dynamics of the fibronectin-III domains of *Aplysia californica* cell adhesion molecules

Catherine M. Kelly,[ab] Julien Muzard,[c‡] Bernard R. Brooks,[d] Gil U. Lee[c] and Nicolae-Viorel Buchete[ab]*§

Received (in XXX, XXX) Xth XXXXXXXXX 2015, Accepted Xth XXXXXXXXX 2015
DOI: 10.1039/b000000x

Due to their homophilic and heterophilic binding properties, cell adhesion molecules (CAMs) such as integrin, cadherin and the immunoglobulin superfamily CAMs are of primary importance in cell-cell and cell-substrate interactions, signalling pathways and other crucial biological processes. We study the molecular structures and conformational dynamics of the two fibronectin type III (Fn-III) extracellular domains of the *Aplysia californica* CAM (apCAM) protein, by constructing and probing an atomically-detailed structural model based on apCAM's homology with other CAMs. The stability and dynamic properties of the Fn-III domains, individually and in tandem, are probed and analysed using all-atom explicit-solvent molecular dynamics (MD) simulations and normal mode analysis of their corresponding elastic network models. The refined structural model of the Fn-III tandem of apCAM reveals a specific pattern of amino acid interactions that controls the stability of the β-sheet rich structure and could affect the apCAM response to physical or chemical changes of its environment. It also exposes the important role of several specific charged residues in modulating the structural properties of the linker segment connecting the two Fn-III domains, as well as of the inter-domain interface.

Introduction

Cell adhesion molecules (CAMs) are glycoproteins located on cell surfaces, and are involved in cell-cell signalling and cell-substrate recognition. They are divided into four main families: integrins, selectins, cadherins and the immunoglobulin superfamily CAMs (IgCAMs). CAMs can be essential therapeutic targets for the treatment of cancer and neurodegenerative diseases.[1-5]

The subject of our study is apCAM (*Aplysia californica* cell adhesion molecule), an IgCAM found on the neuronal growth cones (see Fig. 1A) of *Aplysia californica*, a species of sea slug also known as the California sea hare[6, 7]. apCAM is involved in neurite outgrowth and fasciculation, the mediation of growth cone steering and synaptic plasticity[6, 8-11]. Due to the large size of its neurons and its simple nervous system architecture, *Aplysia californica* is an important model organism for studying certain molecular and cellular mechanisms underlying learning and memory, which would be much more challenging to investigate in the more complex mammalian system. A number of key molecules and important molecular pathways have been discovered first in *Aplysia californica*[12]. For example, a main signalling pathway initiated by apCAM is the activation of the mitogen-activated protein kinase (MAPK) cascade[13].

Like many IgCAMs, apCAM is composed of three parts: cytoplasmic, transmembrane and extracellular. When apCAM's extracellular region (see Fig. 1B) takes part in adhesion, it leads directly to signal transduction. The signal is transmitted through the membrane to the cytoplasmic section, which interacts with the cytoskeleton, triggering a cellular response (*e.g.* neurite outgrowth)[14]. As illustrated in Fig. 1B, the extracellular part of apCAM consists of five immunoglobulin-like (Ig) domains (Ig1-5) followed by two fibronectin type III (Fn-III) domains (Fn1 and Fn2) – the focus of this paper – located near the cell membrane.

Mayford *et al.*[15] used three independent cDNA clones encoding three related apCAM isoforms of varying sequence lengths - 765 residues (UniProt Q9BKQ1), 812 residues (UniProt Q9BKQ0), and 932 residues (UniProt Q9BKP9). All three isoforms share an identical 734-residue sequence at the amino terminus, beginning with a typical 27-residue hydrophobic signal peptide sequence[16, 17]. In the 812-residue and 932-residue isoforms, but not in the 765-residue isoform, this 734-residue sequence is followed by a glutamate-rich acidic region. The two shortest isoforms do not include transmembrane or cytoplasmic regions. They each terminate in a hydrophobic residue sequence common to proteins which are anchored to the membrane by a glycosylphosphoinositol (GPI) linkage. In the longest 932-residue isoform (studied here), a 30-residue highly hydrophobic transmembrane region and a 117-residue cytoplasmic region are present at the C-terminus. BLAST searches for these transmembrane and cytoplasmic sequences identify several high-scoring alignments with other cell adhesion molecules, particularly for NCAM (over a wide range of species, including human NCAM).

A primary role of apCAM is to mediate specific homophilic adhesion processes by interacting with other apCAMs located on adjacent neurons (as illustrated schematically in Fig. 1A)[18]. Mathematical models describing the corresponding association

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/pccp

PAPER

Table 1 MD simulation parameters of the three atomistic systems of apCAM fibronectin type III domains (Fn1, Fn2 and the entire, connected Fn1-Fn2 tandem), solvated with explicit TIP3P water molecules.

System	Simulation time (ns)	No. of residues	Water molecules	Total no. of atoms	Initial dimensions (Å)
Fn1 monomer	50	94	4,875	16,008	60×47×61
Fn2 monomer	50	114 (w. linker)	6,029	19,884	56×59×65
Fn1-Fn2 tandem	250	208	15,931	50,970	72×74×102

and dissociation kinetics for apCAM-apCAM contacts have indicated that these homo-interactions involve two binding modes – one involving all five immunoglobulin domains, and another involving only Ig1 and Ig2 (distal from the cell membrane)[18, 19].

To date, no high-resolution experimental structural data has been made available for apCAM. In order to understand the structural context for apCAM-apCAM interactions, we model its structure using its mammalian homologue, neural CAM (NCAM), as a template.[20-26] X-ray crystal structures are available for individual domains and two-domain complexes of human NCAM (*e.g.* Fn1-Fn2 (PDB ID 2VKW), Ig5-Fn1 (PDB ID 3MTR), Fn1 (PDB ID 2HAZ))[27, 28]. NCAM and apCAM are known to have functional and structural similarity[29-31] and a relatively low but significant sequence identity[32].

Due to the large size of the apCAM protein, it is not practical to run long, all-atom MD simulations on the entire molecule. We focus initially on the fibronectin type III tandem (Fn1-Fn2) located near the C-terminus, which is distinct from the remaining Ig-like domains near the N-terminus.

Fibronectin is a large extracellular matrix protein involved in cell adhesion. It is composed of repeats of three modules; types I, II and III. Each module has a β-sandwich structure formed by two anti-parallel β-sheets surrounding a hydrophobic core. The Fn-III domain is one of the most common structural motifs found in extracellular proteins, often acting as a structural spacer[33, 34]. Notably, Fn-III modules with low sequence identities (*i.e.* smaller than 20%) can show very high structural similarity[35].

Fn-III domains may partially unfold when force is applied, due to the absence of the disulphide bonds which stabilise Fn-I and Fn-II domains[36]. This may explain the elastic properties exhibited by fibronectin fibrils. The elasticity inherent in Fn-III modules may play a vital role in apCAM-apCAM adhesion, allowing apCAM to stretch reversibly without detaching from the cell surface. This extensibility is not present in the Ig domains, which in CAMs almost universally possess a disulphide bond between the pair of cysteines in each Ig fold, limiting the degree of extension.[37-39]

Non-equilibrium states for Fn-III domains appear to play an important part in the function of fibronectin. Stretching of Fn-III modules allows the exposure of cryptic binding sites[40-42]. The mechanical unfolding pathways of Fn-III domains have been investigated in a number of previous studies using a variety of molecular dynamics techniques, including steered molecular dynamics (SMD) simulations, to explore the structure and mechanical stability crucial to key functions of the protein. Several different X-ray and NMR-derived Fn-III structures were used, including those found in human NCAM and in titin.[43] The results are in good agreement with experimental studies carried out using single-molecule force spectroscopy techniques[44, 45].

In apCAM, the Fn-III tandem is located near the membrane surface (Fig. 1). It may function as a spacer, allowing the more adhesive Ig domains to extend further and explore a larger area in the search for a binding partner. This, combined with the extensibility of Fn-III discussed above, points toward the Fn-III tandem playing a crucial role in the apCAM-apCAM adhesion mechanism.

Methods

In this work, we use a combination of homology modelling, large-scale all-atom molecular dynamics (MD) simulations and coarse-grained normal mode analysis (NMA) methods[20-23] to generate and test a structural model of the two fibronectin type III domains of the apCAM ectodomain: Fn1 and Fn2 respectively. Based on the sequence identity of these domains to the corresponding NCAM domains, we build, validate, refine and analyse atomic-resolution models for Fn1 and Fn2 using a variety of molecular modelling approaches, from simplified models of elastic networks to all-atom MD simulations with explicit water molecules[32, 46-52]. Our resulting all-atom structural models for Fn1 and Fn2 offer unique insights into the structure of the extracellular apCAM domains and represent a first step towards the modelling of the full ectodomain and ultimately of the entire apCAM structure.

Homology modelling

We used the 932 amino acid apCAM sequence available from NCBI (GenBank M89648 to M89650) and from SwissProt[53] (Uniprot Q9BKP9). Subdomains were first identified following the study of Mayford *et al.*[15], and confirmed using the NCBI–Genbank web tools[54]. Searches for homologous proteins were carried out using the FASTA3 program[55] of the European Bioinformatics Institute web server, and also the UniProt[56] and RCSB PDB[57] databanks. Conserved domains were identified using the NCBI conserved motif tool[58] and verified individually. The ClustalW2 program at EBI[59] was used for sequence alignment. A homology model of the entire apCAM ectodomain was built using the comparative protein modelling environment, SWISS-MODEL[60-63]. The PDB structure 2VKW for the fibronectin type III tandem of human NCAM was used as a

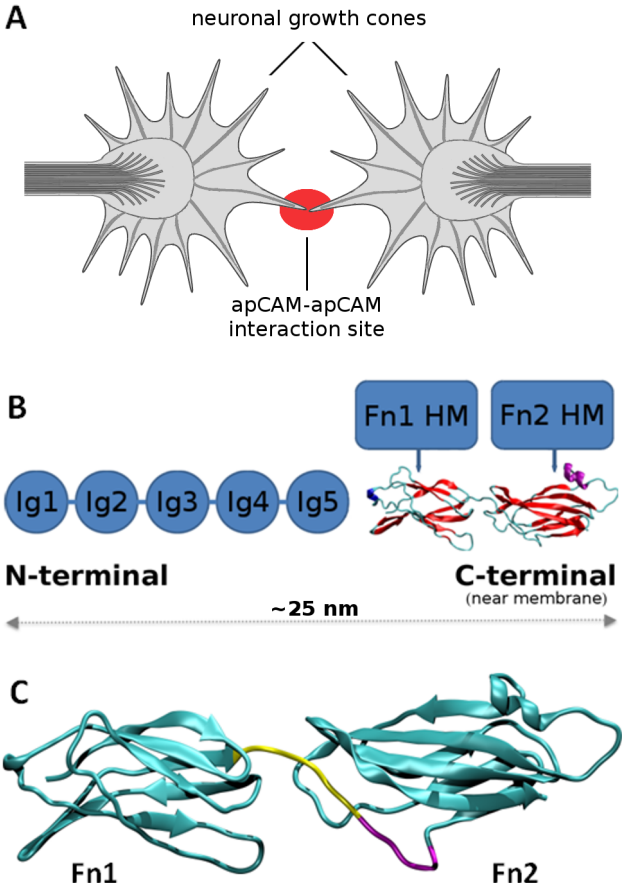

Fig. 1 Schematic illustrations of apCAM domains. (A) Schematic representation of *Aplysia californica* neuronal growth cones, illustrating the location of apCAM-apCAM molecular interactions in the synaptic region. (B) The extracellular apCAM region consists of two fibronectin type III domains (Fn1 and Fn2) proximal to the cellular membrane, and of five immunoglobulin-like domains (Ig1-Ig5). (C) Initial molecular models of the Fn1 and Fn2 domains constructed by homology modelling. The linker region is shown in yellow and magenta. In our HM model, magenta residues are considered to be structurally a part of Fn2.

template,[27] with PDB 2V5Y (receptor protein tyrosine phosphatase mu, RPTPμ, another IgCAM) also used as a template for sections of the molecule not covered by existing NCAM structures.[64]

All-atom validation and refinement using MD methods

Three systems were prepared for atomistic MD simulation using the homology model structures for Fn1 and Fn2: the individual Fn1 and Fn2 domains, and a connected Fn1-Fn2 tandem (see Table 1 for simulation parameters). In the models used for the separated monomer simulations, the linker region between the domains was connected to Fn2. When preparing the system for simulation, normal protonation states at physiological pH were assumed. The single histidine residue was protonated at the epsilon position. Each system was solvated with explicit TIP3P water molecules[65] prior to minimisation, heating and equilibration. Total atom numbers for each system, including water molecules, are reported in Table 1.

MD simulations were performed with the NAMD2 software[66] using the CHARMM22 force field with the CMAP correction.[67, 68] All simulations were performed in the NPT ensemble (*i.e.* constant number of atoms, pressure and temperature), using periodic boundary conditions. We used the modified Nosé-Hoover Langevin piston method implemented in NAMD[69, 70] with a damping time of 0.1 ps, while maintaining a pressure of 1.01325 bar. The temperature was set to 310 K and controlled using a Langevin thermostat with a damping coefficient of $10ps^{-1}$. The Particle Mesh Ewald method was used to include electrostatic effects[71]. The switching distance for non-bonded electrostatics and van der Waals interactions was 9.5 Å with a cutoff distance of 12 Å. The integration time step was 1 fs.

VMD[72] and PyMOL[73] were used for analysis and visualisation. Protein secondary structure was calculated using STRIDE.[74]

Normal mode analysis using ENM

The AD-ENM (Analysis of Dynamics of Elastic Network Model) server[47] was used to perform analysis of our refined structure. An elastic network model was built by connecting nearby α-carbon atoms with elastic springs of uniform spring constant. A distance cutoff of 10 Å was chosen. Normal mode analysis was performed to yield a spectrum of normal modes for the elastic network model, and the zero modes corresponding to translations and rotations (modes 0 to 5) were excluded. A range of frequencies was obtained for the normal modes, and the ability of the model to deform elastically was calculated. The low-frequency end of the spectrum is of particular interest because the lowest modes are able to capture collective conformational changes that are difficult to access through all-atom molecular dynamics simulations.

Normalised squared mode fluctuations were calculated using the WEBnm@ server.[75]

Results and discussion

Structural model for fibronectin type III domains

apCAM molecules are located on the surface of neuronal growth cones in *Aplysia californica*. Fig. 1A schematically illustrates the interaction site of two apCAMs on adjacent growth cones. The molecular details behind this complex adhesion mechanism are as yet not fully understood.[9, 15, 18, 19, 32]

As illustrated in Fig. 1B, apCAM's ectodomain consists of five immunoglobulin-like domains (labelled Ig1 to Ig5), and two fibronectin type III domains (labelled Fn1 and Fn2). The ectodomain is part of an apCAM-mediated cell-cell adhesion molecular system corresponding to an intercellular distance of ~30-40 nm, similar to NCAM-NCAM intercellular distances[76]. The arrangement of the five Ig and two Fn domains in apCAM is similar to that observed for other glycoproteins expressed on the surface of developing axons, such as NCAM in vertebrates, and fasciclin II in invertebrates. These molecules mediate cell-cell adhesion as well as axon growth and guidance.[9, 15, 18, 28]

In this paper, we focus only on the fibronectin type III tandem of apCAM adjacent to the cell membrane. Due to its sheer size, it is not currently feasible to simulate the entire 7-domain extracellular apCAM with atomistic resolution. The Fn1-Fn2 domains are structurally (and possibly functionally) distinct from the five neighbouring Ig-like domains which are distal from the membrane.

The conformation of the linker amino acids between the two Fn domains (Fig. 1C) could greatly affect the position of the Ig domains involved in adhesion. The behaviour of the domains

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/pccp

PAPER

Fig. 2 Sequence comparison between Fn1 and Fn2 domains of apCAM and human NCAM. Sequence alignments of apCAM and human NCAM Fn1 and Fn2 domains and their corresponding percent identity and percent similarity values. Here, Fn1 and Fn2 are illustrated as assigned in Mayford *et al.*[15]. Three binding motifs, involved in FGFR1 (FRM and FGL) and ATP binding, are labelled in the black boxes (see Fig. 9). The yellow linker region shown in Fig. 1C is also highlighted in yellow. The apCAM residue numbers are indicated above the alignment (grey), and the secondary structure of human NCAM is shown below. The colours along the sequence correspond to high consensus (red), low consensus (blue), and neutral (black).

close to the membrane can also have a significant effect on the rest of the molecule.

Mayford *et al.* first published the sequence of apCAM in 1992[15], with their model assigning residues GLY526-SER631 to Fn1 and residues PRO632-ALA736 to Fn2. However, our initial, homology-based model indicates that Fn1 covers the range GLY526-THR618, and that Fn2 covers the range ARG626-SER733. Based on their structural flexibility, confirmed by subsequent MD modelling (see below) we assign the seven residues LEU619-PHE625 to a distinct linker segment between the Fn-III domains. In our model, residues ARG626-SER631 are in contact with the Fn2 domain (see Fig. 1C). These residues are solvent-exposed, which may have led to some ambiguity in the original domain assignment by Mayford *et al.*[15], which placed a domain boundary between residues SER631 and PRO632.

The sequence alignments presented in Fig. 2 do not include the transmembrane and cytoplasmic regions of apCAM and NCAM. Although the amino acid sequences of the transmembrane and cytoplasmic regions are available from Mayford *et al.*[15], there was no suitable template upon which to model an initial structure. The amino acid sequence of the extracellular portion of the molecule (Ig1, Ig2, Ig3, Ig4, Ig5, Fn1, Fn2) was used for our subsequent modelling.

In our models, both Fn1 and Fn2 domains adopt the *β*-sandwich-type fold typical of all Fn-III domains, consisting of seven strands arranged in two anti-parallel sheets (see Fig. 1C). Based on our sequence alignment (see Fig. 2) and homology modelling, the Fn1 and Fn2 domains of our apCAM structural model exhibit a spatial arrangement similar to the corresponding Fn domains of human NCAM, though these domains have relatively low sequence identities of 23.6% and 29.4%, respectively. Although Fn-III modules may have high structural similarity despite relatively low sequence identities, we perform additional tests to ensure that the refined structure is conformationally stable. We subject the homology-based model to extensive atomistic molecular dynamics simulations using

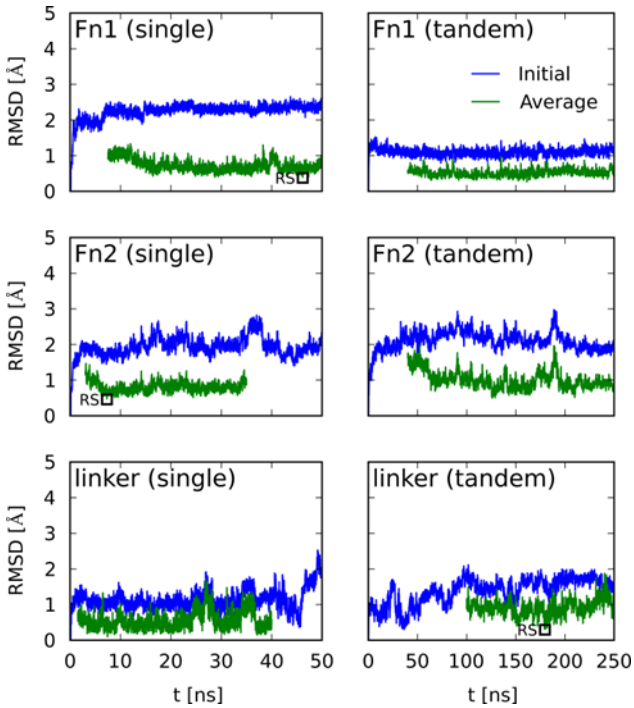

Fig. 3 MD-based structural refinement of homology models for Fn1 and Fn2 domains. RMSD values during individual simulations of the separated monomers Fn1 and Fn2 were calculated with respect to the initial (blue) and average (green) domain structures. The corresponding RMSD values for the same domains when part of a connected tandem are also shown. Representative structures for the Fn1 (46.17 ns) and the Fn2 (7.29 ns) monomers are indicated. A representative structure for the tandem (180.00 ns) is chosen based on the RMSD fluctuations of the linker region (residues 619-625).

explicit water solvation, both as monomers and as a connected tandem, as described below. A visual comparison of the NCAM template and our initial homology-based apCAM model is provided in the supplementary information (Fig. S1).

All-atom MD results

We use trajectories from atomistic MD simulations for both the individual Fn1 and Fn2 domains, and for the entire connected Fn1-Fn2 tandem, to extract representative structures for our model (see Fig. 3). The MD simulation parameters used for these three systems are summarised in Table 1.

As in our previous work[77] we select representative structures from a trajectory by calculating the RMSD with respect to the average structure along the converged portion of the trajectory, $RMSD_{avg}$ (see green line in Fig. 3), and then choosing the frame with the lowest $RMSD_{avg}$ value.[72, 77] We calculate RMSDs for a core subset of residues incorporating the majority of the system, particularly the stable β-sheet regions, while excluding a small number of residues showing higher flexibility. Using this method, we are able to demonstrate the convergence of our structures, which would otherwise be hidden in standard RMSD analysis due to the constant movement of certain flexible loops in the protein.[77]

We consider a system to have achieved convergence once the RMSD with respect to the initial structure (see blue line in Fig. 3) plateaus. The individually, fully solvated Fn1 and Fn2 monomer systems (see Table 1) achieve convergence in our MD simulations after approximately 15 ns and 6 ns, respectively (see Fig. 3). While Fn2 fluctuates more than Fn1 due to its disordered coil content near the untethered ends, they both maintain relatively stable structures thereafter.

In the Fn1-Fn2 tandem trajectory, Fn1 reaches a clear convergence after about 10 ns and maintains an RMSD of approximately 1 Å from the initial homology model structure with no significant fluctuation. Fn2 also reaches convergence but again, shows a wider fluctuation around the average RMSD of approximately 2 Å (see Fig. 3). SI Movies S1 and S2 illustrate clearly the equilibrium structural fluctuations of Fn1 and Fn2, respectively.

Based on our RMSD analysis, as explained above, we choose a representative structure for the Fn1 domain corresponding to a simulation time of 46.17 ns, and a representative structure for the Fn2 domain corresponding to simulation time of 7.29 ns (indicated in Fig. 3). Since the conformation of the linker is an important factor in determining the overall structure of the tandem, we choose the representative structure for the tandem based on the RMSD values calculated for the linker. This method gives us a representative structure corresponding to a simulation time of 180 ns.

The secondary structure predicted by the homology model is reasonably well-conserved (see Figs. 4 and 5), although there is a loss of β-content for each domain. Note the relatively small decrease in β-content for Fn1, contrasting with the relatively large decrease in β-content for Fn2 due to disorder in the loops at the C-terminal. This is due to the lack of restraint placed upon the ends of the Fn1-Fn2 tandem in our simulations, allowing freedom of movement which would not occur when the tandem is situated within the entire apCAM protein. We therefore restrict our structural analysis primarily to the linker and interface regions of the tandem.

The initial homology models and the corresponding representative structures for each monomer and for the tandem are compared in Fig. 5. For the short monomer simulations, the

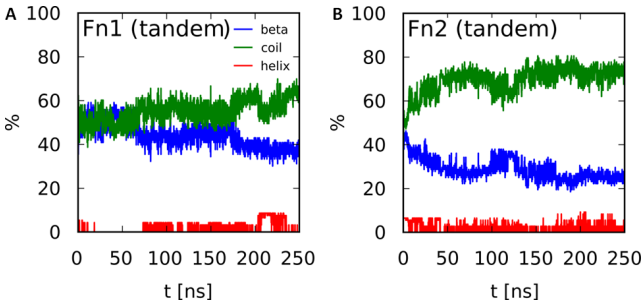

Fig. 4 Secondary structure of the Fn1-Fn2 tandem. Secondary structure types (i.e. β-strand, coil and helical) shown as percentage of overall structure for (A) Fn1 and (B) Fn2 domains during the Fn1-Fn2 tandem simulation.

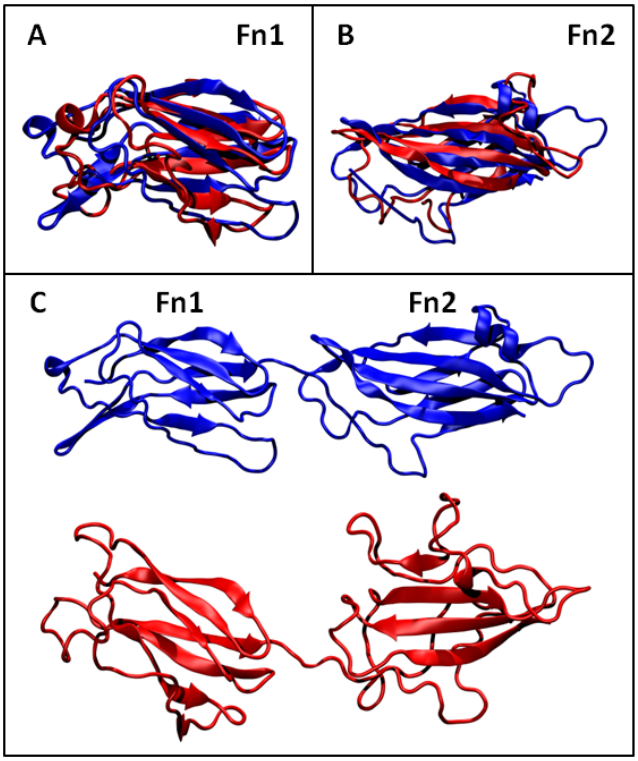

Fig. 5 Comparison of homology-based structures and MD-refined representative structures for Fn1, Fn2 and the Fn1-Fn2 tandem. Homology-based structures (blue) and MD-refined representative structures (red) are shown for (A) the Fn1 monomer, (B) the Fn2 monomer, and (C) the Fn1-Fn2 tandem. For clarity, the tandem structures are shown next to each other rather than being overlaid.

representative structures align very closely with the homology models, although there is some minor rearrangement of loops and coil content. The representative tandem structure has a more marked contrast, particularly in Fn2 where significant β-content has been lost. This can be at explained by the fact that the C-terminal of Fn2 is not anchored, allowing for more mobility at this end. This is also true of the N-terminal of Fn1, but the β-strand located here is more stable and resistant to unravelling. There is also a slight change in the angle between the two domains. However, on the whole, the representative structure for the tandem is remarkably similar to the homology model.

The essentially electrostatic-controlled interactions at the Fn1-Fn2 interface are illustrated in Fig. 6. These play a vital role in stabilising the structure of each domain and also of the entire

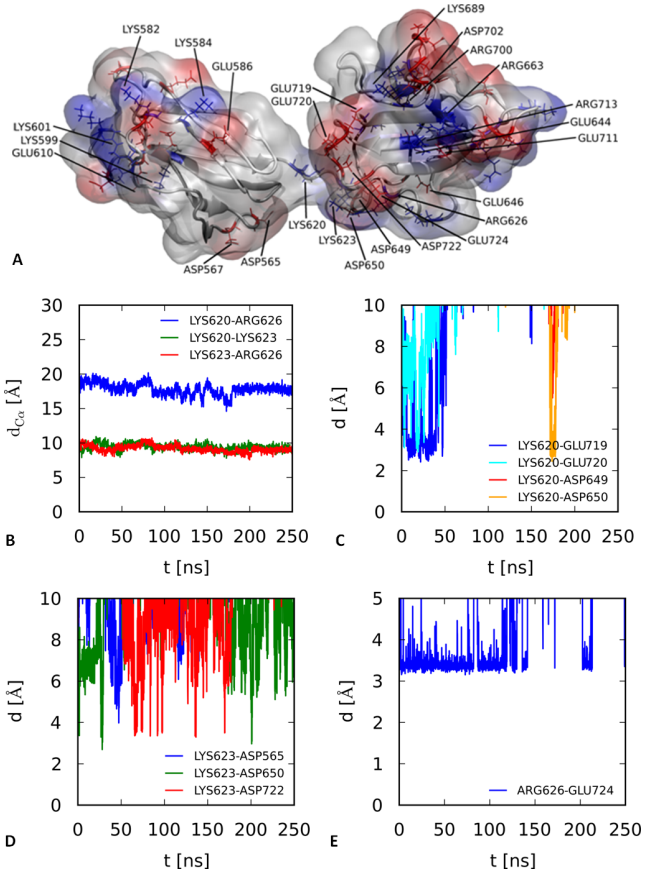

Fig. 6 Electrostatic interactions at the Fn1-Fn2 interface. (A) Charged residues in the Fn1-Fn2 tandem. Negatively-charged amino acids are shown as red, the positive ones as blue, and neutral as white. Charged residues are labelled with their residue ID. (B) Distances between the C_α atoms of LYS620, LYS623 and ARG626, illustrating the constant extension of the backbone along the linker between the two Fn-III domains. (C, D, and E) Distances between the terminal heavy atoms of charged linker residues - LYS620 (C), LYS623 (D), ARG626 (E) - and their nearby salt bridge interaction partners.

tandem. Charged residues are highlighted in Fig. 6A.

Three positively-charged residues are situated along the linker (LYS620, LYS623, ARG626), as shown in Fig. 6A. Mutual repulsion maintains a steady distance between each of these positive charges, keeping the linker extended throughout the entire trajectory (see Fig. 6B).

The presence of the corresponding negatively-charged residues in an adjacent area of Fn2 (GLU719, GLU720, ASP722, GLU724) allows salt bridges to form and to maintain the orientation of Fn2 with respect to the linker (see Figs. 6C-6E).

For the first ~50 ns of the Fn1-Fn2 tandem trajectory, LYS620 forms a stable salt bridge with GLU719, occasionally switching to the adjacent GLU720 (see Fig. 6C). In the later stages of the simulation, this interaction ceases as the loop containing GLU719 and GLU720 moves away from the linker and folds back toward Fn2. GLU719 forms salt bridges with LYS689 and ARG713 in the main body of Fn2, and this loop remains locked in place for the remainder of the trajectory. At this point, with its former binding partners rendered inaccessible, LYS620 rotates downward. This position is stabilised by the brief formation of salt bridges with ASP649 and ASP650, on the lower surface of

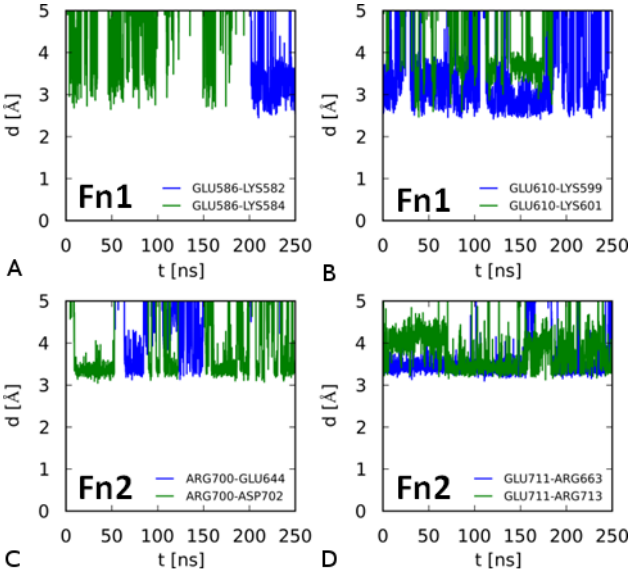

Fig. 7 Electrostatic interactions within Fn1 and Fn2. Competing salt bridge formation involving sets of three residues (e.g., GLU586, LYS582 & LYS584 in Fn1) that remain in close proximity for the duration of the trajectory.

Fn2 toward the interface (see Fig. 6C).

In the initial conformation of the tandem, LYS623 is located close to ASP650 on the surface of Fn2, and the two residues briefly form a salt bridge. Over the course of ~60 ns the Fn2 loop in which ASP722 is located approaches the linker, allowing a salt bridge to form intermittently with LYS623 over the next ~110 ns. This is the primary electrostatic interaction that LYS623 experiences. At approximately 175 ns, the Fn2 loop moves away from the linker, breaking the bond, and LYS623 forms a brief salt bridge once more with ASP650 (see Fig. 6D).

A stable salt bridge forms between ARG626 and GLU724 for the majority of the trajectory (see Fig. 6E) helping to keep Fn2 locked in a relatively stable position.

In addition, the negatively charged residues located on the lower linker-facing surface of each domain (ASP565, ASP567 on Fn1, ASP650, ASP649, GLU646 on Fn2) further reinforce the separation of the two domains. Note that the gap between the negatively charged Asp residues on either side of the Fn1-Fn2 interface is never less than ~5 Å in our simulations (see Fig. S2C in the supplementary information), and is significantly higher for most of the remaining trajectory.

In the model studied here, the conformation of the tandem appears thus to be controlled primarily by electrostatic interactions, resulting in an extended linker and a clear separation of the Fn1 and Fn2 domains, while allowing enough flexibility that the linker is not trapped in a conformational state that would be too rigid for proper signal transduction.

Electrostatic interactions also play a role in stabilising the Fn1 and Fn2 domains themselves. Here, we see several interesting three-residue salt bridge interactions, where two residues of a certain charge compete to bind with a nearby residue of the opposite charge. The result is that a certain charged residue (*e.g.* GLU586 in Fn1 or ARG700 in Fn2) binds competitively either with one or the other of its oppositely charged partners at all times, causing it to remain locked in position. This competition is illustrated clearly in Fig. 7C, where, for example, the positively

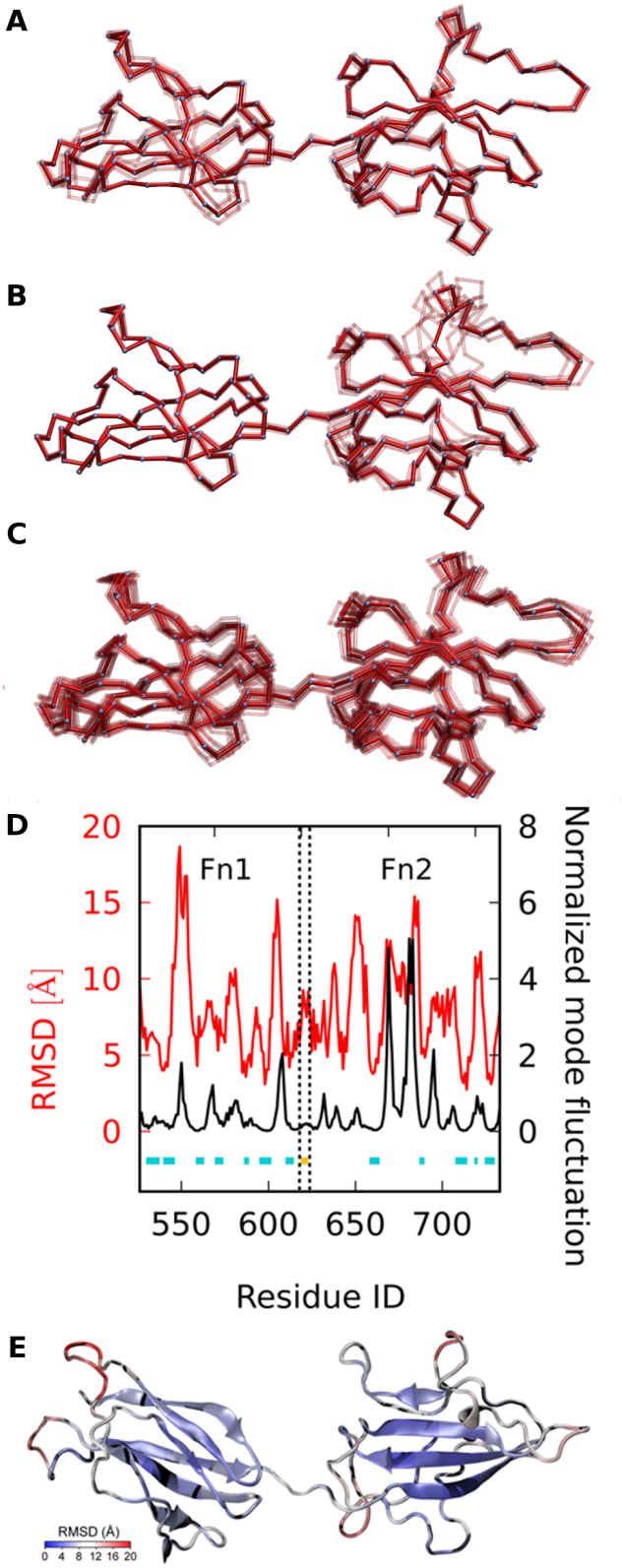

Fig. 8 Normalised mode fluctuation and localised movement in the Fn-III tandem. Minimum and maximum deformation of the first ten non-zero modes (modes 6 to 15) as calculated for the representative structure of the tandem by normal mode analysis. (A) Modes 6 and 10 mostly describe motion in Fn1. (B) Modes 9, 13 and 15 mostly describe motion in Fn2. (C) Modes 7, 8, 11, 12 and 14 describe general movement of the tandem. Each mode is shown separately in the supplementary material – see Figs. S3 and S4. (D) Normalised mode fluctuation (average of the first six non-zero modes) and RMSD by residue. Secondary structure is shown above the x-axis; β-sheets in cyan, linker in yellow. (E) Fn1-Fn2 tandem coloured according to average RMSD across the trajectory of the MD simulation. Areas shown in blue exhibit little movement; areas shown in red are more mobile.

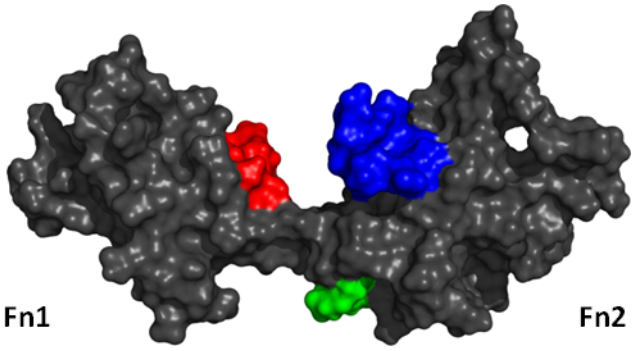

Fig. 9 MD-refined representative structure of the Fn1-Fn2 tandem. Surface representation of the Fn1-Fn2 tandem with motifs involved in ATP-binding (green) and FGFR1-binding (FRM in red; FGL in blue) highlighted.

charged ARG700 alternates between forming a salt bridge with GLU644 and with ASP702.

Future work could provide additional quantitative free energy calculations of the inter-domain interactions, which could be based on classical MD or corrected by using more accurate quantum mechanics/molecular mechanics (QM/MM) calculations.[78, 79] Additional analysis of the inter-domain interface shows that other non-covalent contacts may also be important. In particular, while the Fn1 domain has a clear hydrophobic patch in the vicinity of amino acids ILE616, PRO617, LEU619, VAL535, and LEU536, the Fn2 domain presents a much smaller hydrophobic surface due to amino acids PRO654 and ILE655. Future availability of longer trajectories may allow additional quantitative analysis (e.g., by using programs such as NCIPlot).[80, 81]

The ENM normal modes of Fn1 and Fn2

Normal mode analysis was performed on the elastic network model of the MD-refined tandem structure. Modes corresponding to translations and rotations were excluded, and the fluctuation of the ten lowest non-zero modes (modes 6-15) is shown visually in Fig. 8A-C. Normal modes in the lowest frequencies tend to describe the largest motions in a protein, and hence are the most functionally relevant. Modes 6 and 10 mostly describe motion in Fn1 (Fig. 8A). Modes 9, 13 and 15 mostly describe motion in Fn2 (Fig. 8B). The remaining modes, 7, 8, 11, 12 and 14, describe general movement of the tandem and do not seem to be linked to any major localised changes (Fig. 8C). Each mode is shown separately in the supplementary material (Figs. S3 and S4). Unsurprisingly, there is little movement in the β-sheet regions and higher movement in the loops and disordered coil areas. This implies that the β-sheets are quite rigid, while there is more flexibility in the loops and coil.

The mode fluctuation for the first six non-zero modes is shown in a more quantitative manner in Fig. 8D, where it is compared to the actual motion exhibited by each residue throughout the simulation trajectory.

As expected, the structural regions that show significant conformational change throughout the MD simulation are shown by the normal mode analysis to have higher flexibility. There is

good agreement between the two plots, with β-strands (cyan) remaining relatively fixed, and loop and coil regions showing more freedom of movement. This is further demonstrated in Fig. 8E, where each residue of the Fn1-Fn2 tandem is coloured according to average RMSD across the trajectory. The correlation coefficient for the two quantities is calculated to be 0.43. When the residues corresponding to sharp peaks in the normalised mode fluctuation between residue 668 and 683 are excluded, the correlation coefficient increases to 0.50, indicating a moderate correlation between mode fluctuation and RMSD by residue.

Potential binding sites located on the Fn1-Fn2 tandem

Interestingly, several binding motifs observed in NCAM are highly conserved in apCAM (see Fig. 2). One example is the DDGG binding motif associated with ATP hydrolysis in NCAM (highlighted in green in Fig. 9). The locations of two motifs (FRM, red, and FGL, blue) involved in FGFR1 binding are also shown in Fig. 9. This observation points towards a possibly common functional relevance of these motifs.

Conclusions

We study the structural features and conformational dynamics of the fibronectin type III domains, Fn1 and Fn2, of the extracellular segment of the apCAM protein. Based on homology modelling primarily using the human NCAM crystal structures (PDB ID 2VKW), we build initial structural candidate models for individual domains, Fn1, Fn2, and their covalently-connected tandem, Fn1-Fn2. Similarly to the corresponding Fn tandem in the human NCAM protein, the proposed initial apCAM Fn tandem structures are predominantly composed of β-strand rich domains connected by a linker segment. We update the initial domain assignment of Mayford *et al.*[15] based on the structure of our new homology model, and we define a 7-residue linker region between the two domains.

Our equilibrated model of the Fn1-Fn2 tandem reveals an interesting distribution of positively-charged amino acids along the inter-domain linker, and the corresponding negatively-charged residues found in the Fn2 structure, which contribute strongly to the stability of the entire Fn1-Fn2 apCAM segment. Using information from our large-scale all-atom MD simulations with explicit water molecules, we find a molecular mechanism in which specific strong electrostatic interactions between charged residues stabilise the Fn1-Fn2 tandem, both at the interface between the two domains and within the Fn1 and Fn2 domains themselves. In particular, residues LYS620, LYS623 and ARG626 on the linker appear to play a crucial dual role, as revealed by our MD simulations: (i) they facilitate an extended conformation of the linker region and, at the same time, (ii) they play a stabilising role for the Fn2 domain via salt bridges with residues GLU719, GLU720, ASP722 and GLU724. These findings may be probed directly by introducing mutations of these positively-charged linker residues and running further all-atom MD simulations to investigate the stability of the structure when the charges of these key residues are altered and the major electrostatic interactions between the linker and the neighbouring regions are interrupted.

Using long atomistic MD equilibration simulations, we obtain stable structures for both the individual Fn1 and Fn2 domains (see SI Movie S1 and S2 for visualising the equilibrium structural fluctuations of Fn1 and Fn2, respectively) and, importantly, for the model of the connected Fn1-Fn2 tandem. Our study of the normal modes corresponding to the representative tandem structure, together with per-residue RMSD values calculated using the entire 250 ns tandem trajectory, shows high stability for the β-sheet segment and significant mobility in the connected loops which may have functional roles in the protein-protein recognition and interaction processes involving apCAM.

Interestingly, our model exhibits ATP and FGFR1 (FRM and FGL) binding sites similar to those found in the NCAM Fn-III tandem (Figs. 2 and 9).[32, 35, 40, 82] This observation raises the possibility of signalling-related modulation of apCAM functionality, suggesting a convergent evolution of homologous CAMs such as apCAM and NCAM that preserved structural features important for binding and signalling.

Our study opens the possibility of constructing larger atomistically-detailed models for the entire 7-domain extracellular part of apCAM[15], and possibly of other cell adhesion molecules. More importantly, our Fn1-Fn2 apCAM structural model provides the first piece in the puzzle leading to a detailed atomistic model for the highly specific apCAM-apCAM interactions, and could advance the understanding of the complex molecular mechanisms involved in cell adhesion processes.[18, 19, 32, 35, 40]

Supplementary Information

The supplementary information includes four figures (Fig. S1: Comparison of Fn-III domain structures for the initial homology-based apCAM model and for the NCAM crystal structure template, Fig. S2: Structural stability of the entire Fn1-Fn2 tandem along the 250 ns atomistic MD simulation, Figs. S3 and S4: Deformation for the lowest ten non-zero modes), and two movies showing the equilibrium structural fluctuations of the Fn1 and Fn2 domains in apCAM's Fn1-Fn2 tandem.

Acknowledgements

The authors thank the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC), and the Biowulf cluster at the National Institutes of Health, USA (http://biowulf.nih.gov) for the provision of computational facilities and support. CMK and NVB would like to thank Frank McQuillan for useful initial discussions.

This work was supported in part by the University College Dublin Structured Ph.D. Programme in Simulation Science, funded under the Programme for Research in Third Level Institutions (PRTLI) Cycle 5 and co-funded by the European Regional Development Fund (ERDF).

Notes

[a] *School of Physics, University College Dublin, Belfield, Dublin 4, Ireland. Fax: +353-(0)1-283-7275; Tel: +353-(0)1-716-2088; E-mail: buchete@ucd.ie*

[b] *Complex and Adaptive Systems Laboratory (CASL), University College Dublin, Belfield, Dublin 4, Ireland.*

[c] *UCD Nanomedicine Centre, School of Chemistry and Chemical Biology, University College Dublin, Belfield, Dublin 4, Ireland.*
[d] *Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.*
† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/
‡ Current address: The Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA.
§ CMK, JM, GUL and NVB designed the study. JM constructed the initial homology model. CMK performed the MD simulations and normal mode analysis. CMK, BRB and NVB analysed the MD data. All authors discussed the results and the manuscript. CMK and NVB wrote the manuscript.

References

1 M. A. Bewick and R. M. Lafrenie, *Curr. Pharm. Des.*, 2006, **12**, 2833--2848.
2 L. H. Brennaman and P. F. Maness, in *Structure and Function of the Neural Cell Adhesion Molecule NCAM*, ed. V. Berezin, Springer-Verlag New York, 2010, pp. 299-317.
3 G. F. Le Bras, K. J. Taubenslag and C. D. Andl, *Cell Adh. Migr.*, 2012, **6**, 365-373.
4 S. A. Mousa, *Mol. Biotechnol.*, 2007, **38**, 33--40.
5 X. Zhong and F. J. Rescorla, *Cell. Signal.*, 2012, **24**, 393--401.
6 F. Keller and S. Schacher, *J. Cell Biol.*, 1990, **111**, 2637--2650.
7 C. Thompson, C. H. Lin and P. Forscher, *J Cell Sci.*, 1996, **109**, 2843--2854.
8 D. M. Suter and P. Forscher, *J. Neurobiol.*, 2000, **44**, 97--113.
9 C. O. Mejean, A. W. Schaefer, K. B. Buck, H. Kress, A. Shundrovsky, J. W. Merrill, E. R. Dufresne and P. Forscher, *PLOS ONE*, 2013, **8**.
10 D. M. Suter, L. D. Errante, V. Belotserkovsky and P. Forscher, *J. Cell Biol.*, 1998, **141**, 227--240.
11 D. M. Suter and P. Forscher, *Curr. Opin. Neurobiol.*, 1998, **8**, 106--116.
12 E. R. Kandel, M. Brunelli, J. Byrne and V. Castellucci, *Cold Spring Harbor Symp. Quant. Biol.*, 1976, **40**, 465-482.
13 C. H. Bailey, B. K. Kaang, M. Chen, K. C. Martin, C. S. Lim, A. Casadio and E. R. Kandel, *Neuron*, 1997, **18**, 913-924.
14 P. F. Maness and M. Schachner, *Nat. Neurosci.*, 2007, **10**, 19--26.
15 M. Mayford, A. Barzilai, F. Keller, S. Schacher and E. R. Kandel, *Science*, 1992, **256**, 638-644.
16 G. Von Heijne, *Nucleic Acids. Res.*, 1986, **14**, 4683-4690.
17 G. von Heijne, *Eur. J. Biochem.*, 1983, **133**, 17-21.
18 E. Martines, J. Zhong, J. Muzard, A. C. Lee, B. B. Akhremitchev, D. M. Suter and G. U. Lee, *Biophys. J.*, 2012, **103**, 649-657.
19 D. Kilinc, A. Blasiak, J. J. O'Mahony, D. M. Suter and G. U. Lee, *Biophys. J.*, 2012, **103**, 1120-1129.
20 C. Chothia and A. M. Lesk, *EMBO J.*, 1986, **5**, 823-826.
21 P. G. Jambrina, O. Bohuszewicz, N.-V. Buchete, W. Kolch and E. Rosta, *Biochem. Soc. Trans.*, 2014, **42**, 784-790.
22 T. Schwede, *Structure*, 2013, **21**, 1531-1540.
23 C. L. Worth, S. Gong and T. L. Blundell, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 709-720.
24 Y. Zhang, *Curr. Opin. Struct. Biol.*, 2009, **19**, 145-155.
25 A. Kryshtafovych, K. Fidelis and J. Moult, *Proteins: Struct., Funct., Bioinf.*, 2014, **82**, 164-174.
26 M. Compiani and E. Capriotti, *Biochemistry*, 2013, **52**, 8601-8624.
27 F. Carafoli, J. L. Saffell and E. Hohenester, *J. Mol. Biol.*, 2008, **377**, 524-534.
28 C. Kasper, H. Rasmussen, J. S. Kastrup, S. Ikemizu, E. Y. Jones, V. Berezin, E. Bock and I. K. Larsen, *Nat. Struct. Biol.*, 2000, **7**, 389--393.
29 R. Kleene, M. Mzoughi, G. Joshi, I. Kalus, U. Bormann, C. Schulze, M.-F. Xiao, A. Dityatev and M. Schachner, *J. Neurosci.*, 2010, **30**, 10784--10798.
30 H. Cremer, G. Chazal, C. Goridis and A. Represa, *Mol. Cell Neurosci.*, 1997, **8**, 323--335.
31 P. Doherty, M. S. Fazeli and F. S. Walsh, *J. Neurobiol.*, 1995, **26**, 437--446.
32 V. Maruthamuthu, K. Schulten and D. Leckband, *Biophys. J.*, 2009, **96**, 3005-3014.
33 I. D. Campbell and C. Spitzfaden, *Structure*, 1994, **2**, 333--337.
34 A. H. Huber, Y. E. Wang, A. J. Bieber and P. J. Bjorkman, *Neuron*, 1994, **12**, 717--731.
35 D. Craig, M. Gao, K. Schulten and V. Vogel, *Structure*, 2004, **12**, 21-30.
36 H. P. Erickson, *J. Muscle Res. Cell. Motil.*, 2002, **23**, 575-580c.
37 H. P. Erickson, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 10114-10118.
38 A. Soteriou, A. Clarke, S. Martin and J. Trinick, *Proc. Biol. Sci.*, 1993, **254**, 83-86.
39 P. Carl, C. H. Hwok, G. Manderson, D. W. Speicher and D. E. Discher, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 1565-1570.
40 M. Gao, D. Craig, O. Lequin, I. D. Campbell, V. Vogel and K. Schulten, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 14784-14789.
41 I. Vakonakis, D. Staunton, L. M. Rooney and I. D. Campbell, *EMBO J.*, 2007, **26**, 2575-2583.
42 C. Zhong, M. Chrzanowska-Wodnicka, J. Brown, A. Shaub, A. M. Belkin and K. Burridge, *J. Cell. Biol.*, 1998, **141**, 539-551.
43 D. Craig, A. Krammer, K. Schulten and V. Vogel, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 5590-5595.
44 N. I. Abu-Lail, T. Ohashi, R. L. Clark, H. P. Erickson and S. Zauscher, *Matrix Biol.*, 2006, **25**, 175-184.
45 H. Li, H. H.-L. Huang, C. L. Badilla and J. M. Fernandez, *J. Mol. Biol.*, 2005, **345**, 817-826.
46 C. M. Kelly, T. Northey, K. Ryan, B. R. Brooks, A. L. Kholkin, B. J. Rodriguez and N.-V. Buchete, *Biophys. Chem.*, 2015, **196**, 16-24.
47 W. J. Zheng and S. Doniach, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 13253-13258.
48 N. V. Buchete, J. E. Straub and D. Thirumalai, *Polymer*, 2004, **45**, 597-608.
49 F. Tofoleanu, B. R. Brooks and N. V. Buchete, *ACS Chem. Neurosci.*, 2015.
50 N. V. Buchete, *Biophys. J.*, 2012, **103**, 1411-1413.
51 A. L. Milac, N. V. Buchete, T. A. Fritz, G. Hummer and L. A. Tabak, *J. Mol. Biol.*, 2007, **373**, 439-451.
52 N. V. Buchete, J. E. Straub and D. Thirumalai, *Proteins: Struct., Funct., Bioinf.*, 2008, **70**, 119-130.
53 B. Boeckmann, *Nucleic Acids. Res.*, 2003, **31**, 365-370.
54 D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and D. L. Wheeler, *Nucleic Acids. Res.*, 2008, **36**, D25-D30.
55 W. Pearson, *Methods Mol. Biol.*, 1992, **132**, 195-219.
56 A. Bairoch, R. Apweiler, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi and L. S. Yeh, *Nucleic Acids. Res.*, 2005, **33 (suppl. 1)**, D154-D159.
57 A. Kouranov, L. Xie, J. de la Cruz, L. Chen, J. Westbrook, P. E. Bourne and H. M. Berman, *Nucleic Acids. Res.*, 2006, **34 (suppl. 1)**, D302-D305.
58 A. Marchler-Bauer, C. Zheng, F. Chitsaz, M. K. Derbyshire, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, S. Lu, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang and S. H. Bryant, *Nucleic Acids. Res.*, 2013, **41**, D348-352.
59 M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, *Bioinformatics*, 2007, **23**, 2947-2948.
60 M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. Gallo Cassarino, M. Bertoni, L. Bordoli and T. Schwede, *Nucleic Acids. Res.*, 2014, **42**, W252-W258.
61 K. Arnold, L. Bordoli, J. Kopp and T. Schwede, *Bioinformatics*, 2006, **22**, 195-201.
62 F. Kiefer, K. Arnold, M. Künzli, L. Bordoli and T. Schwede, *Nucleic Acids. Res.*, 2009, **37**, D387-D392.
63 N. Guex, M. C. Peitsch and T. Schwede, *Electrophoresis*, 2009, **30**, S162-S173.

64 A. R. Aricescu, C. Siebold, K. Choudhuri, V. T. Chang, W. Lu, S. J. Davis, P. A. van der Merwe and E. Y. Jones, *Science*, 2007, **317**, 1217-1220.
65 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926--935.
66 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781-1802.
67 B. R. Brooks, C. L. Brooks III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545-1615.
68 A. D. Mackerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400-1415.
69 G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**.
70 S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, 1995, **103**.
71 T. A. Darden, D. M. York and L. G. Pederson, *J. Chem. Phys.*, 1993, **98**, 10089.
72 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33-38.
73 Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.3r1, 2010.
74 M. Heinig and D. Frishman, *Nucleic Acids. Res.*, 2004, **32**, W500-502.
75 S. M. Hollup, G. Saelensminde and N. Reuter, *BMC Bioinformatics*, 2005, **6**, 52.
76 C. P. Johnson, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 6963-6968.
77 F. Tofoleanu and N. V. Buchete, *J. Mol. Biol.*, 2012, **421**, 572--586.
78 G. König and B. R. Brooks, *Biochim. Biophys. Acta, Gen. Subj.*, 2014.
79 E. Rosta, M. Klähn and A. Warshel, *J. Phys. Chem. B*, 2006, **110**, 2934-2941.
80 J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan and W. Yang, *J. Chem. Theory Comput.*, 2011, **7**, 625-632.
81 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498-6506.
82 M. Gao, D. Craig, V. Vogel and K. Schulten, *J. Mol. Biol.*, 2002, **323**, 939-950.

